*Proc. Natl. Acad. Sci. USA*
Vol. 88, pp. 9438–9442, November 1991
Biochemistry

# Structure of the human hexabrachion (tenascin) gene

**(type III units/fibrinogen/epidermal growth factor repeats)**

JEFFREY R. GULCHER, DONALD E. NIES, MARK J. ALEXAKOS, NEEJU A. RAVIKANT, MICHAEL E. STURGILL, LINDA S. MARTON, AND KARI STEFANSSON*

Departments of Neurology and Pathology (Neuropathology) and the Committees on Neurobiology and Immunology, The University of Chicago, Chicago, IL 60637

*Communicated by Hewson Swift, July 22, 1991 (received for review June 25, 1991)*

**ABSTRACT** **The structure of the gene encoding human hexabrachion (tenascin) has been determined from overlapping clones isolated from a human genomic bacteriophage library. The genomic inserts were characterized by restriction mapping, Southern blot analysis, PCR, and DNA sequencing. The coding region of the hexabrachion gene spans ≈80 kilobases of DNA and consists of 27 exons separated by 26 introns. The exon–intron structure supports a hypothesis based on the cDNA sequence that the hexabrachion gene is an assembly of DNA modules that are also found elsewhere in the genome. Single exons may encode a module, a portion of a module, or a group of modules. The 15 type III units similar to those found in fibronectin are each encoded either by a single exon or by two exons interrupted by an intron. All type III units known to be spliced out of the smaller forms of the protein are encoded by one exon. The fibrinogen-like domain of 210 amino acids is encoded by five exons. The 14.5 epidermal growth factor-like repeats are all encoded by a single exon.**

Hexabrachion (Hxb) (also called tenascin and cytotactin) is a large glycoprotein that appears in extracellular matrices as a disulfide-linked multimer (1). It is expressed in large amounts at specific times and locations during development and is found in small amounts in certain adult tissues including the central nervous system, soft tissue stroma, cartilage, and regions of basement membrane (1). Hxb may be reexpressed to significant levels in tumor matrices and in areas of wound healing (2, 3). Rotary-shadowed electron micrographs of Hxb show a protein with six flexible arms associated with each other at one end such that they radiate outward from the center of the complex and terminate in a globular domain (4). The polypeptides are all derived from a single gene that can produce several isoforms through alternative splicing of a primary transcript (5–8). The human Hxb gene has been localized to chromosome 9q32–34 (9).

We have reported (10) sequences of two forms of human Hxb deduced from cDNA (GenBank accession no. M55618). The larger form is 2203 amino acids and the smaller one is 1573 amino acids. At the N terminus, there is a group of hydrophobic residues and eight cysteine residues believed responsible for the association of six monomers into a homohexamer (7, 8). C-terminal to this region are 14.5 epidermal growth factor (EGF)-like repeats, which are similar to the domains found in the precursor of the EGF and a growing number of proteins such as the low density lipoprotein receptor and laminin (11, 12). Immediately adjacent to the EGF-like repeats are 15 similar but nonidentical ≈90-amino acid units. These contiguous modules show sequence similarity to the type III units found in fibronectin and several other proteins (13, 14). All known splicing events involving Hxb RNAs occur through variable retention or exclusion of members of this group of type III units. The difference between the two forms we have reported (5) lies in segments of RNA encoding seven type III units (units 6–12) that are spliced out of the mRNA for the smaller form but are retained in the mRNA encoding the larger one. Type III units have been described in a bacterial protein, indicating that the type III units are old structures on an evolutionary scale and may have appeared prior to the emergence of eukaryotes (15). The C terminus of the protein contains a 210-amino acid domain rich in basic residues that is similar to the C termini of the $\beta$ and $\gamma$ chains of human fibrinogen (7, 16). This report describes the intron–exon organization of the coding portion of the human Hxb gene.

## METHODS

**Isolation and Characterization of Hxb Genomic Clones.** Five unamplified genomic libraries representing 12–15 million recombinant phage were prepared from DNA isolated from peripheral blood mononuclear cells of a single donor (17). The libraries were initially screened with probes derived from human Hxb cDNA clones (16) (Fig. 1) and subsequently with genomic clones. Initially, 26 phage clones were identified, plaque purified, and scaled-up for characterization. The λ DNA was first analyzed by restriction mapping. Select clones were then digested with *Eco*RI and *Eco*RI/*Sal* I. The digests were subcloned directly with "shotgun" ligation (41) into Bluescript plasmid (Stratagene) prepared by digestion with *Eco*RI and *Eco*RI/*Sal* I. Isolated subclones were further analyzed by restriction mapping, on Southern blots, by the PCR, and by DNA sequencing (17–19).

**Determination of Intron–Exon Structure.** The ends of each subclone were first sequenced. Two clones from λ3 and -44 (Fig. 1) were found to contain portions of type III unit 11 and the second exon of the fibrinogen domain, respectively, at their ends. In addition, exons VII, XI, XII, XVII, and XXII were localized by sequencing the ends of subclones from λ81, -85, -16, and -18. With these seven subclones serving as a scaffold, the remaining subclones were further analyzed by restriction mapping based on a fine restriction map of the cDNA. Restriction deletions of select subclones were prepared and parts of all other exons were localized with the exception of exon X. The location of exon X was determined using Southern blots and the PCR (18). Finally, this exon was sequenced using specific primers that annealed to known sequences within the cDNA.

## RESULTS

The coding region of the human Hxb gene was found to consist of a total of 27 exons separated by 26 introns (Fig. 1).

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: Hxb, hexabrachion; EGF, epidermal growth factor.

*To whom reprint requests should be addressed at: Department of Neurology, BH Box 425, The University of Chicago, 5841 South Maryland Avenue, Chicago, IL 60637.

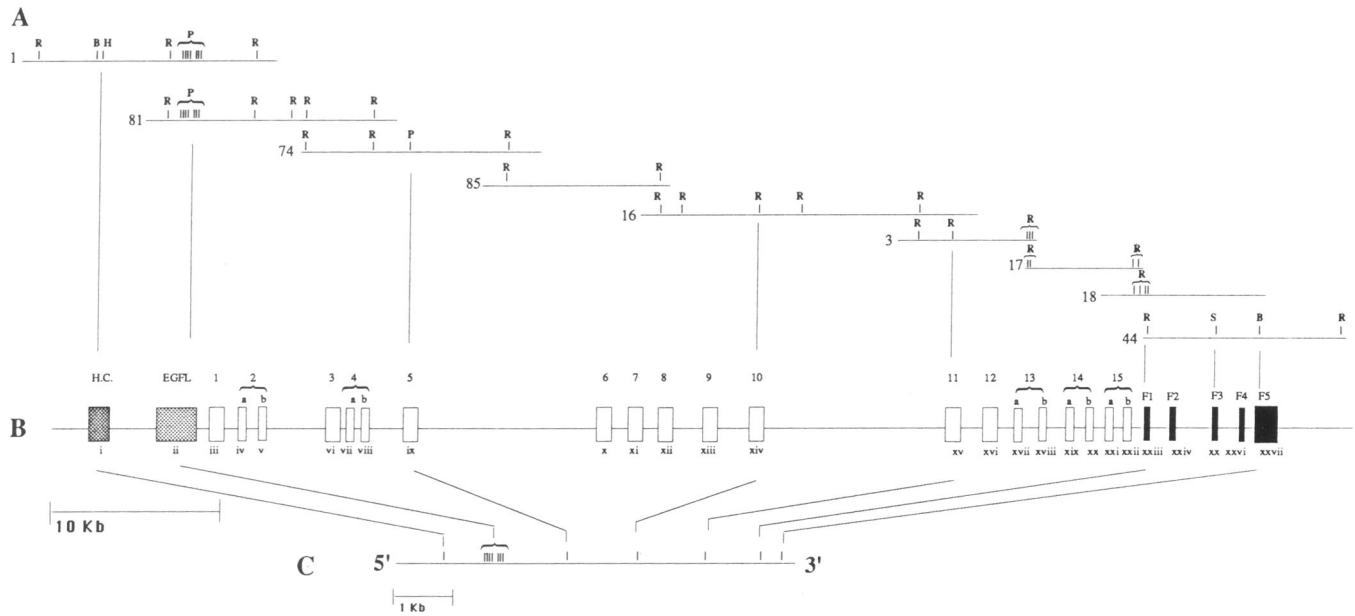


FIG. 1. (*A*) Genomic clones used to reconstruct the exon–intron structure of the gene. Vertical lines represent restriction sites for *Eco*RI (R), *Pst* I (P), *Hin*dIII (H), and *Sac* I (S). (*B*) Exon–intron structure. Horizontal lines represent introns and boxes are exons. The scale applies only to the sizes of the introns; the exons are smaller than the scale indicates (the actual sizes of the exons are listed in Table 1). Shaded boxes are exons encoding 5′ untranslated region of the transcript, the cysteine-rich N-terminal domain, and EGF-like repeats. Open boxes are exons encoding the type III units and solid boxes are exons encoding the fibrinogen-like domain and the 3′ untranslated region of the transcript. Roman numerals refer to the numbers of the exons (5′–3′), whereas Arabic numerals refer to the type III units. (*C*) cDNA derived from the largest Hxb transcript we have characterized. Vertical lines represent restriction sites that correspond to some of the restriction sites in the genomic clones.

All exons were completely sequenced and corresponded with few exceptions† to the published cDNA sequence, and all intron–exon junctions conformed to the GT–AG rule. The sequences of all intron–exon junctions and the sizes of each exon are shown in Table 1. Exon I encodes the first 153 amino acids, encompassing the translational start site, the leader sequence, and six of the eight N-terminal cysteines that constitute the cysteine-rich region believed responsible for holding the Hxb monomers together in a hexavalent complex. Recently published data show that the chicken Hxb gene contains an additional 5′ exon encoding an untranslated segment of the mRNA solely (20). This upstream exon is separated from the exon containing the translational start site by a 12-kilobase (kb) intron (20).

The first exon is separated from the second exon by a 3-kb intron (Fig. 1). The second exon spanning 1410 base pairs encodes 20 amino acids of the cysteine-rich region including two of the cysteines and all 14.5 tandem repeats of 31-amino acids that show a high degree of similarity to the peptide hormone EGF and to a family of related EGF-like repeats found in a diverse group of proteins (21).

†Sequence discrepancies: The genomic sequences of the exons were compared to the complete cDNA sequence (10) and were helpful in resolving the few discrepancies between the amino acid sequences determined by our group and by Siri *et al.* (8). We found an extra residue, glycine (residue 244), between the two alanine residues (residues 243 and 244 in ref. 8) and the genomic sequence confirms our finding. We found leucine instead of valine at residue 370, glutamine instead of arginine at residue 680, and leucine instead of isoleucine at residue 1677 (residues 369, 679, and 1676, respectively, in ref. 8) and in each case the genomic sequence confirms our sequence. However, the genomic sequence confirms their suspicion that our residues 1600–1608 were incorrect because of two frame-shifts, and we agree that the correct sequence is Ser-Gly-Phe-Thr-Gln-Gly-His-Gln-Thr (residues 1599–1607 in ref. 8). The genomic sequence also confirms their claim that our residue 1066 is arginine rather than histidine. The genomic sequence confirms residues 2142–2145, Tyr-Lys-Gly-Ala, which differs from the Thr-Arg-Gly found by Siri *et al.* (8).

The human Hxb gene encodes at least 15 repeats equivalent to the type III units found in fibronectin that vary in size between 87 and 91 amino acids. It is within the region of the primary transcript encoding the type III units where alternative splicing occurs to produce distinct isoforms of human (and chicken and mouse) Hxb (6–9, 22). The segment encoding the first type III unit is separated from the exon containing the EGF-like repeats by a 1-kb intron. Each type III unit is encoded by one or two exons. For type III units 1, 3, and 5–12, each unit is encoded by a single exon and, for units 2, 4, and 13–15, each unit is encoded by two exons. The location of the introns interrupting the segments encoding the latter units is not the same for any two of the five introns. However, these introns split the region encoding the central third of the type III units, which is less conserved than the ends. The positions of the introns in the five "split" type III units are shown in Fig. 2. The type III units that are known to be spliced out of the smaller forms of the protein are each encoded by one exon. The introns separating exons encoding the type III units range in size from 600 base pairs to 10 kb and it is possible that one or more of the regions we have labeled as introns contain exons that are retained in mRNAs for forms that have yet to be characterized.

The fibrinogen-like domain is at the distal end of the Hxb arms and is most similar to the C-terminal halves of the $\beta$ and $\gamma$ chains of fibrinogen. The region of the gene encoding the fibrinogen-like domain is separated from the second exon of the 15th type III unit by a 1-kb intron, and it is encoded by a total of five exons. The three human fibrinogen genes each contain several introns whose locations are in one case conserved in two genes and in two cases conserved in all three genes (16). However, of the four introns splitting the region encoding the fibrinogen-like domain of Hxb, one is located at a site that coincides with an intron in the $\beta$-chain gene of fibrinogen, whereas the remaining three introns do not correspond to introns in the fibrinogen genes. The final exon of the fibrinogen-like domain is 753 nucleotides long and encodes the last 36 amino acids before the stop codon, and it

Table 1. Sequences of the exon–intron junctions of the Hxb gene

| Domain | 5′-Side intron splice acceptor | | | | | | | Exon number | 3′-Side intron splice donor | | | | | | | Exon size, bp |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| H.C. | | | | | | | | I | CCT P | GCC A | ACA T | G G | gta | tga | gct | >600 |
| EGFL | ttg | ttg | cag | GC | CGC R | TTG L | GAC D | II | TGC C | TCA S | GAG E | G V | gtg | agt | gca | 1410 |
| III-1 | tcc | tgc | cag | TG | TCT S | CCT P | CCC P | III | GTG V | GCC A | ACG T | T Y | gtg | agt | gag | 264 |
| III-2a | tac | tta | cag | AC | TTA L | CCT P | GCA A | IV | CGG R | AAT N | ATG M | | gta | agg | aga | 116 |
| III-2b | atc | tct | cag | AAT N | AAA K | GAA E | GAT D | V | ACC T | ACC T | ACA T | C R | gtg | agt | atc | 157 |
| III-3 | cca | tca | tag | GC | TTG L | GAT D | GCC A | VI | TTC F | ACA T | ACA T | G G | gta | aaa | ggc | 270 |
| III-4a | ctc | aaa | cag | GC | CTC L | GAT D | GCA A | VII | ACA T | CTC L | ACA T | G G | gtg | agt | ctc | 186 |
| III-4b | tct | cct | tag | GT | CTG L | AGG R | CCG P | VIII | GCA A | GCC A | ACA T | G E | gtg | ggg | cac | 90 |
| III-5 | tcc | tcc | cag | AG | TTG L | GAC D | ACG T | IX | GCA A | TCC S | ACT T | G E | gtg | tgt | ctg | 264 |
| III-6 | cta | ctt | cag | AA | CAA Q | GCC A | CCT P | X | GCC A | TCC S | ACA T | G G | gta | tct | ctt | 273 |
| III-7 | ctt | ctc | cag | GG | GAA E | ACT T | CCC P | XI | GTC V | TTG L | ACA T | G E | gta | ttc | tag | 273 |
| III-8 | ctt | atg | cag | AG | GAG E | GTT V | CCA P | XII | GTC V | GTC V | ACA T | G E | gta | tga | ccc | 273 |
| III-9 | ctt | ctt | cag | AG | GAT D | CTC L | CCA P | XIII | GCC A | TCC S | ACA T | G A | gta | ctt | cct | 273 |
| III-10 | ctt | tta | cag | CC | AAA K | GAA E | CCT P | XIV | GCC A | ACG T | ACA T | G E | gta | cat | gtg | 273 |
| III-11 | ttc | ttt | cag | AG | GCC A | CTG L | CCC P | XV | ATT I | GTT V | ACA T | G E | gta | ttt | caa | 273 |
| III-12 | tcc | ctc | tag | AA | GCC A | GAA E | CCG P | XVI | GCA A | ACA T | ACA T | G A | gta | ctg | taa | 273 |
| III-13a | ttc | tga | cag | CC | ATG M | GGC G | TCC S | XVII | ATT I | ACA T | GGA G | G G | gta | gga | gac | 123 |
| III-13b | gtg | ttc | cag | GT | ACA T | CCC P | TCC S | XVIII | TTC F | ACC T | ACA T | G A | gta | gtt | ggt | 144 |
| III-14a | tct | ttg | cag | CT | CTG L | GAT D | GGC G | XIX | GGC G | GAG E | AAA K | G V | gta | agg | ttt | 120 |
| III-14b | cat | gtt | tag | TG | CCA P | GAA E | ATT I | XX | TTC F | ACA T | ACA T | G D | gta | caa | aga | 144 |
| III-15a | ctt | ctc | cag | AC | CTC L | GAT D | TCT S | XXI | ACA T | GTC V | AAG K | | gta | ttt | aag | 132 |
| III-15b | gct | ctt | tag | GAA E | GTC V | ATT I | GTG V | XXII | TTC F | ACC T | ACA T | A I | gta | gga | ggc | 132 |
| F1 | tat | cca | cag | TT | GGA G | CTC L | CTG L | XXIII | GGA G | TGG W | ATT I | | gtg | agt | acc | 152 |
| F2 | tgt | tta | tag | GTG V | TTC F | CTG L | AGA R | XXIV | TTC H | TGG W | CTT L | G G | gta | atg | cag | 103 |
| F3 | cat | cca | cag | GG | CTG L | GAC D | AAC N | XXV | AGT S | GGG G | ACA T | G A | gta | cag | tag | 162 |
| F4 | gtc | tct | cag | CA | GGT G | GAC D | TCC S | XXVI | CAC H | AGT S | CAG Q | | gta | gag | agg | 164 |
| F5 | ttt | taa | cag | GGC G | GTT V | AAC N | TGG W | XXVII | AAA K | CGG R | GCA A | | TAA STOP | | | 753 |

Capital letters are used for nucleotides in exons and lower-case letters are used for nucleotides in introns. H.C., cysteine-rich N-terminal domain; EGFL, domain containing the EGF-like repeats; III, type III unit; F, exons encoding the fibrinogen-like domain; bp, base pairs.

contains a single canonical polyadenylylation signal reported in a full-length cDNA (10).

## DISCUSSION

This report describes the intron–exon structure of the coding portion of the human Hxb gene. The data presented conform to the proposal that Hxb is a modular protein, whose sequence modules can be found in somewhat different but recognizable forms in many functionally diverse proteins. No two types of modules were found to reside on a common exon.

The organization of the part of the gene encoding the type III units, each of which is encoded by one or two exons, makes it easy to envisage its assembly from primordial components by exon shuffling and duplication, utilizing breaks and ligations in intronic regions that, if not functionally inert, may tolerate to a much greater degree disruptions of linear architecture than do exons (23). The introns that separate the regions encoding each type III unit in the Hxb gene are all phase I introns; that is, they separate the first and the second bases of a codon. Phase I introns also separate the regions encoding almost all type III units in fibronectin (24), the neural cell adhesion molecule (25), the growth hormone receptor (26), and the interleukin 2 receptor (27). This arrangement supports the proposal that the type III units in all of these proteins were generated through duplication and reduplications of one primordial unit (14).

However, the EGF-like repeats in Hxb present a somewhat different picture. In the human Hxb gene all 14.5 EGF-like repeats are encoded by a single exon and there is no intronic interruption to define a single repeat. The chicken Hxb

```
1.        SEVSPPKDLVVTEVTEETVNLAWDNEMRVTEYLVVYTPT  HEGGLEM QFRVPGDQTSTIIQELEPGVEYFIRVFAILENKKSIPVSARVAT
2.         YLPAPEGLKFKSIKETSVEVEWDPLDIAFETWEIIFRNMNKEDEGEI TKSLRRPETSYRQTGLAPGQEYEISLHIVKNNTRGPGLKRVTTT
3.         RLDAPSQIEVKDVTDTTALITWFKPLAEIDGIELTYGIKDVPGDRTTID  LTEDENQYSIGNLKPDTEYEVSLISRRGDMSSNPAKETFTT
4.         GLDAPRNLRRVSQTDNSITLEWRNGKAAIDSYRIKYAPISGGDHAEVDVPKSQQATTKTTLTGLRPGTEYGIGVSAVKEDKESNPATINAAT
5.         ELDTPKDLQVSETAETSLTLLWKTPLAKFDRYRLNYSLPTGQWVGVQ    LPRNTTSYVLRGLEPGQEYNVLLTAEKGRHKSKPARVKAST
      abcd
6.     *** EQAPELENLTVTEVGWDGLRLNWTAADQAYEHFIIQVQEANKVEAAR  NLTVPGSLRAVDIPGLKAATPYTVSIYGVIQGYRTPVLSAEAST
7.     *** GETPNLGEVVVAEVGWDALKLNWTAPEGAYEYFFIQVQEADTVEAAQ  NLTVPGGLRSTDLPGLKAATHYTITIRGVTQDFSTTPLSVEVLT
8.     *** EEVPDMGNLTVTEVSWDALRLNWTTPDGTYDQFTIQVQEADQVEEAH  NLTVPGSLRSMEIPGLRAGTPYTVTLHGEVRGHSTRPLAVEVVT
9.     *** EDLPQLGDLAVSEVGWDGLRLNWTAADNAYEHFVIQVQEVNKVEAAQ  NLTLPGSLRAVDIPGLEAATPYRVSIYGVIRGYRTPVLSAEAST

10.     ** AKEPEIGNLNVSDITPESFNLSWMATDGIFETFTIEIIDSNRLLETV  EYNISGAERTAHTSGLPPSTDFIVYLSGLAPSIRTKTISATATT
11.   **** EALPLLENLTISDINPYGFTVSWMASENAFDSFLVTVVDSGKLLDPQ  EFTLSGTQRKLELRGLITGIGYEVMVSGFTQGHQTKPLRAEIVT
12.      * EAEPEVDNLLVSDATPDGFRLSWTADEGVFDNFVLKIRDTKKQSEPL  EITLLAPERTRDLTGLREATEYEIELYGISKGRRSQTVSAIATT
13.         AMGSPKEVIFSDITENSATVSWRAPTAQVESFRITYVPITGGTPSM   VTVDGTKTQTRLVKLIPGVEYLVSIIAMKGFEESEPVSGSFTT
14.         ALDGPSGLVTANITDSEALARWQPAIATVDSYVISYTGEKVPEIT    RTVSGNTVEYALTDLEPATEYTLRIFAEKGPQKSSTITAKFTT
15.         DLDSPRDLTATEVQSETALLTWRPPRASVTGYLLVYESVDGTVKE    VIVGPDTTSYSLADLSPSTHYTAKIQALNGPLRSNMIQTIFTT
```

FIG. 2. Type III units of Hxb aligned with respect to sequence similarity. Arrowheads point to the placements of introns dividing the regions encoding type III units 2, 4, 13, 14, and 15. The a, b, c, and d to the left represent the four smaller forms of Hxb. *, Type III units spliced out of the corresponding isoform.

contains 13.5 EGF-like repeats. Thus since the divergence of the avian and the mammalian lineages either the mammalian Hxb has gained an EGF-like repeat or the avian Hxb has lost one. The probability of having a duplication of exactly one repeat or loss of exactly one repeat is considered to be much greater if it is encoded by a distinct exon flanked by introns (23). Hence, it is likely that, at the time of divergence of the avian and the mammalian lineages, there was at least one distinct exon encoding one EGF repeat allowing for the acquisition of an additional EGF-like repeat by the mammalian Hxb gene through duplication or deletion of one of the EGF-like repeats from the avian gene. The duplication or the deletion would then have been followed by loss of the introns.

Individual EGF-like repeats found in other proteins are encoded by separate exons: each of the eight EGF-like repeats in the EGF precursor protein (28) and each of the seven EGF-like repeats in the low density lipoprotein receptor is encoded by one exon (11). The organization of the EGF-like repeats in the human Hxb gene more closely resembles that of the 10.5 EGF-like repeats in the *lin12* locus of *Caenorhabditis elegans*; 8.5 repeats are encoded by the first exon and the remaining two repeats are equally divided between two other exons (29).

The fibronectin gene is similar to the Hxb gene in that it encodes type III units and yields several isoforms through alternative splicing of the region of the primary transcript encoding the type III units (30, 31). Furthermore, in the fibronectin gene the two type III units that are variably included in mRNA are encoded by single exons, whereas all the invariant units are encoded by two exons (24, 32, 33). Based on the known isoforms of human Hxb, it is not clear whether Hxb adheres to this pattern. We have sequenced two forms of human Hxb (5, 10) and we have evidence for the existence of several other forms based on Western and Northern blot analyses. In addition, Siri *et al.* (8) have shown, through the sequencing of cDNA clones and PCR products, that several additional splicing variants of Hxb mRNA exist that differ in the region flanked by type III units 5 and 13. They found Hxb RNA species that lacked units 6–11, units 6–9 and 11, and only unit 11. In this report we show that, in fact for units 1, 3, and 5–12, each unit is encoded by a single exon. Each of the remaining type III units is encoded by two exons. Fig. 2 summarizes the intron locations within the 15 type III units and indicates which units are spliced out of the known forms of Hxb.

In addition to Hxb and fibronectin, there are several other proteins that contain type III units including the cell-adhesion molecules L1 (34), neural cell adhesion molecule (25), and TAG-1 (35); the extracellular matrix proteins contactin/F11 (36, 37) and type VI collagen $\alpha 3$ (38); and the growth hormone/prolactin receptor family that includes the interleukin 2 receptor, interleukin 4 receptor, granulocyte colony-stimulating factor receptor, the erythropoietin receptor, and the interleukin 6 receptor (14, 26, 27, 39, 40). The structures of the genes encoding the neural cell adhesion molecule, the growth hormone receptor, and the interleukin 2 receptor are known and each of their type III units is encoded by two exons. Furthermore, there are no known forms of these molecules that have one or more type III units spliced out. Thus the Hxb, fibronectin, the neural cell adhesion molecule, the growth hormone receptor, and interleukin 2 receptor genes contain 27 invariant type III units, and for 23 of the 27 units, each unit is encoded by two exons. In addition, these genes encode nine alternatively spliced type III units, each of which is encoded by a single exon. Thus there appears to be a correlation between the exon structure of the type III units found in a variety of different molecules and their use as alternatively spliced domains.

There are no published reports of Hxb variants that exclude the single-exon type III units 1, 3, or 5 nor have we observed such variants. These type III units provide the only exceptions to the rule that type III units encoded by two exons are never spliced out and the type III units encoded by one exon are spliced out of at least one form of the molecule. However, it is possible that variants missing these units exist at specific stages in development or in tissues we have yet to examine. This is a particularly interesting possibility for the type III unit 3 that contains the Arg-Gly-Asp (RGD) sequence that may grant the protein affinity for integrins. A Hxb isoform without type III unit 3 would not bind to the integrin and, therefore, splicing out this unit would be a mechanism to modulate integrin-mediated function of Hxb. It is of interest here that the type III unit 10 of fibronectin, which contains the RGD sequence, is encoded by two exons and is according to our understanding likely to be a part of all fibronectin isoforms (26, 27). Alternatively, absence of introns from the regions encoding the type III units 1, 3, and 5 may reflect splicing possibilities that were never utilized or have fallen out of use during evolution.

The contributions of J.R.G. and D.E.N. to the paper were equal. We are grateful to Ji-Hyon Kim for excellent technical assistance. This work was supported in part by grants from the National Multiple Sclerosis Society and from the National Institutes of Health to K.S.

1. Erickson, H. P. & Bourdon, M. A. (1989) *Annu. Rev. Cell Biol.* **5**, 71–92.
2. Chiquet-Ehrismann, R., Mackie, E. J., Pearson, C. A. & Sakakura, T. (1986) *Cell* **47**, 131–139.
3. Lightner, V. A., Gumkowski, F., Bigner, D. D. & Erickson, H. P. (1989) *J. Cell Biol.* **108**, 2483–2493.
4. Erickson, H. P. & Inglesias, J. L. (1984) *Nature (London)* **311**, 267–269.

5. Gulcher, J. R., Nies, D. E., Marton, L. S. & Stefansson, K. (1989) *Proc. Natl. Acad. Sci. USA* **86,** 1588–1592.
6. Jones, F. S., Hoffman, S., Cunningham, B. A. & Edelman, G. M. (1989) *Proc. Natl. Acad. Sci. USA* **86,** 1905–1909.
7. Spring, J., Beck, K. & Chiquet-Ehrismann, R. (1989) *Cell* **59,** 325–334.
8. Siri, A., Carnemolla, B., Saginati, M., Leprini, A., Casari, G., Baralle, F. & Zardi, L. (1991) *Nucleic Acids Res.* **19,** 525–531.
9. Gulcher, J. R., Alexakos, M. J., LeBeau, M. M., Lemons, R. S. & Stefansson, K. (1990) *Genomics* **6,** 616–622.
10. Nies, D. E., Hemesath, T. J., Kim, J.-H., Gulcher, J. R. & Stefansson, K. (1991) *J. Biol. Chem.* **266,** 2818–2823.
11. Sudoff, T. C., Russell, D. W., Goldstein, J. L., Brown, M. S., Sanchez-Pescador, R. & Bell, G. I. (1985) *Science* **228,** 893–895.
12. Sasaki, M., Kato, S., Kohno, K., Martin, G. R. & Yamada, Y. (1987) *Proc. Natl. Acad. Sci. USA* **84,** 935–939.
13. Petersen, T. E., Thorgersen, H. C., Skorstengaard, K., Vibe-Pedersen, K., Sahl, P., Sottrup-Jensen, L. & Magnusson, S. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 137–141.
14. Patthy, L. (1990) *Cell* **61,** 13–14.
15. Watanabe, T., Suzuki, K., Oyanagi, W., Ohnishi, K. & Tanaka, H. (1990) *J. Biol. Chem.* **265,** 15659–15665.
16. Crabtee, G. R., Comeau, C. M., Fowlkes, D. M., Fornace, A. J., Malley, J. D. & Kant, J. A. (1985) *J. Mol. Biol.* **185,** 1–19.
17. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab., Cold Spring Harbor, NY), pp. 270–293.
18. Scharf, S. J., Horn, G. T. & Erlich, H. A. (1986) *Science* **233,** 1076–1078.
19. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 5463–5467.
20. Jones, F. S., Crossin, K. L., Cunningham, B. A. & Edelman, G. M. (1990) *Proc. Natl. Acad. Sci. USA* **87,** 6497–6501.
21. Appella, E., Weber, I. T. & Blas, F. (1988) *FEBS Lett.* **231,** 1–4.
22. Weller, A., Beck, S. & Ekblom, P. (1991) *J. Cell Biol.* **112,** 355–362.
23. Gilbert, W. (1978) *Nature (London)* **271,** 501.
24. Schwarzbauer, J. E., Patel, R. S., Fonda, D. & Hynes, R. O. (1987) *EMBO J.* **6,** 2573–2580.
25. Owens, G. C., Edelman, G. M. & Cunningham, B. A. (1987) *Proc. Natl. Acad. Sci. USA* **84,** 294–298.
26. Godowski, P. J., Leung, D. W., Meacham, L. R., Galgani, J. P., Hellmiss, R., Keret, R., Rotwein, P. S., Parks, J. S., Laron, Z. & Wood, W. I. (1989) *Proc. Natl. Acad. Sci. USA* **86,** 8083–8087.
27. Shibuya, H., Yoneyama, M., Nakamura, Y., Harada, H., Hatakeyama, M., Minamoto, S., Kono, T., Doi, T., White, R. & Taniguchi, T. (1990) *Nucleic Acids Res.* **18,** 3697–3703.
28. Bell, G. I., Fong, N. M., Stempien, M. M., Wormsted, M. A., Caput, D., Ku, L. L., Urdea, M. S., Rall, L. B. & Sanchez-Pescador, R. (1986) *Nucleic Acids Res.* **14,** 8427–8446.
29. Greenwald, I. (1985) *Cell* **43,** 583–590.
30. Odermatt, E., Tamkun, J. W. & Hynes, R. O. (1985) *Proc. Natl. Acad. Sci. USA* **82,** 6571–6575.
31. Kornblihtt, A. R., Umezawa, K., Vibe-Pedersen, K. & Baralle, F. E. (1985) *EMBO J.* **4,** 1755–1759.
32. Vibe-Pedersen, K., Kornblihtt, A. R. & Baralle, F. E. (1984) *EMBO J.* **3,** 2511–2516.
33. Kubomura, S., Obara, M., Karasaki, Y., Taniguchi, H., Gotoh, S., Tsuda, T., Higashi, K., Ohsata, K. & Hirano, H. (1987) *Biochim. Biophys. Acta* **910,** 171–181.
34. Moos, M., Tacke, R., Scherer, H., Teplow, D., Fruh, K. & Schachner, M. (1988) *Nature (London)* **334,** 701–703.
35. Furley, A. J., Morton, S. B., Manalo, D., Karagogeos, D., Dodd, J. & Jessel, T. M. (1990) *Cell* **61,** 157–170.
36. Ranscht, B. (1988) *J. Cell Biol.* **107,** 1561–1572.
37. Brummendorf, T., Wolff, M., Frank, R. & Rathjen, F. G. (1989) *Neuron* **2,** 1351–1361.
38. Bonaldo, P. & Colombatti, A. (1989) *J. Biol. Chem.* **264,** 20235–20239.
39. Boutin, J. M., Jolicoeur, C., Okamura, H., Gagnon, J., Edery, M., Shirota, M., Banville, D., Dusanter-Fourt, I., Djiane, J. & Kelly, P. A. (1988) *Cell* **53,** 69–77.
40. Larsen, A., Davis, T., Curtis, B. M., Gimpel, S., Sims, J. E., Cosman, D., Park, L., Sorensen, E., March, C. J. & Smith, C. A. (1990) *J. Exp. Med.* **172,** 1559–1570.
41. Greene, J.R. & Guarente, L. (1987) *Methods Enzymol.* **55,** 512–522.